Identification and expression of a cDNA from *Daucus carota* encoding a bifunctional aspartokinase-homoserine dehydrogenase¹

Jane M. Weisemann and Benjamin F. Matthews*

United States Department of Agriculture, Agricultural Research Service, Plant Molecular Biology

Laboratory, Beltsville, MD 20705-2350, USA (*author for correspondence)

Received 16 November 1992; accepted in revised form 24 February 1993

Key words: aspartokinase, homoserine dehydrogenase, cDNA, carrot, chloroplast, gene

Abstract

Aspartokinase (EC 2.7.2.4) and homoserine dehydrogenase (EC 1.1.1.3) catalyze steps in the pathway for the synthesis of lysine, threonine, and methionine from aspartate. Homoserine dehydrogenase was purified from carrot (*Daucus carota* L.) cell cultures and portions of it were subjected to amino acid sequencing. Oligonucleotides deduced from the amino acid sequences were used as primers in a polymerase chain reaction to amplify a DNA fragment using DNA derived from carrot cell culture mRNA as template. The amplification product was radiolabelled and used as a probe to identify cDNA clones from libraries derived from carrot cell culture and root RNA. Two overlapping clones were isolated. Together the cDNA clones delineate a 3089 bp long sequence encompassing an open reading frame encoding 921 amino acids, including the mature protein and a long chloroplast transit peptide. The deduced amino acid sequence has high homology with the *Escherichia coli* proteins aspartokinase I-homoserine dehydrogenase I and aspartokinase II-homoserine dehydrogenase II. Like the *E. coli* genes the isolated carrot cDNA appears to encode a bifunctional aspartokinase-homoserine dehydrogenase enzyme.

Abbreviations: AK, aspartokinase; HSDH, homoserine dehydrogenase; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate

Introduction

In plants, the essential amino acids lysine, threonine, methionine and isoleucine are synthesized from aspartate. These amino acids are required in the diets of non-ruminant animals, thus, there is much interest in increasing the quantity of these amino acids in food sources [1]. The first reaction in the pathway, the conversion of aspartate into β -aspartyl phosphate, is catalyzed by the enzyme

The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number L11529.

¹ The mention of vendor or product does not imply that they are endorsed or recommended by the U.S. Department of Agriculture over vendors of similar products not mentioned.

aspartokinase (AK; EC 2.7.2.4). In the second reaction of the pathway, β -aspartyl phosphate is converted to aspartate semialdehyde. At this point the pathway branches: one branch leads to the formation of lysine and the other to the formation of threonine and isoleucine or methionine. The first reaction in this second branch is the reduction of β -aspartyl phosphate to homoserine by the enzyme homoserine dehydrogenase (HSDH; EC 1.1.1.3). These biosynthetic pathways in plants are similar to pathways found in bacteria [4, 5]. Many of the bacterial genes that code for the enzymes of the aspartate pathway have been cloned and sequenced [5]. In contrast the only plant gene coding for an enzyme of the aspartate pathway which has been isolated is that for dihydrodipicolinate synthase, which catalyzes the first reaction specific to lysine synthesis [8].

In most of the plant species studied multiple forms of AK and HSDH have been identified in vitro [1]. These forms are distinguished by their sensitivity to feedback inhibition (in particular by threonine and lysine) and by their molecular weight and subunit composition. From carrot (Daucus carota L.) suspension cultures two forms of HSDH have been identified in vitro: one sensitive and one insensitive to threonine inhibition [13, 21]. These two forms can interconvert. Three forms of AK have been isolated from carrot: form I is strongly inhibited by lysine, form II is strongly inhibited by threonine, and form III is partially inhibited by both [24]. The relationship between these three forms has not yet been defined.

The isolation of clones of these genes from plants should enable us to answer questions about the relationship between various forms of the enzymes, the number of genes involved, and the regulation of the pathway. In *E. coli* there are three genes coding for AK and HSDH [4, 5]. One, *lysC*, codes for a lysine-sensitive AK and is regulated by lysine. The other two genes code for bifunctional AK-HSDH proteins. Expression of the *thrA* gene is repressed by threonine and isoleucine and the enzymatic activity of AKI-HSDHI is inhibited by threonine. Expression of the *metL* gene is repressed by methionine, but

the protein AKII-HSDHII is not responsive to end-product inhibition [4, 5]. It is not known if the multiple enzyme forms in plants are encoded by separate genes or if these genes are subject to transcriptional or translational regulation. Knowledge gained from the study of the aspartate pathway genes may eventually allow the engineering of the pathway to alter the amino acid pool composition of plants used as protein sources.

Our previous biochemical studies of carrot AK and HSDH indicated that these activities resided on the same protein molecule as a bifunctional enzyme [24]. This paper describes the isolation of a carrot cDNA encoding a bifunctional AK-HSDH. In fact, this is the first report of a gene encoding a bifunctional AK-HSDH protein outside of the Enterobacteriaceae.

Materials and methods

General DNA cloning methods

Plasmid DNA preparation, ligations, restriction enzyme digestions, Southern blots, and northern blots were done according to standard procedures [12]. Plant genomic DNA was extracted by the method of Keim et al. [10]. Enzymes were obtained from either Bethesda Research Laboratories or Boehringer Mannheim. PCR reagents were obtained from Perkin-Elmer Cetus. Oligonucleotides used for PCR were obtained from Synthecell.

DNA probes were labelled with $[\alpha^{-32}P]$ dCTP (DuPont-New England Nuclear) using the random oligo priming method [7].

DNA sequence determination was done by the dideoxy chain termination method using modified T7 DNA polymerase (Sequenase 2.0 from United States Biochemical Corp.) [20].

Construction of a carrot cDNA library

Total RNA was extracted from two-month old carrot roots by the method of Chirgwin et al. [3].

Poly(A)⁺ RNA was isolated from total RNA on columns of oligo(dT) cellulose [12], cDNA was made from 3 μ g poly (A)⁺ RNA using the cDNA Synthesis System Plus from Amersham. Eco RI adapters were added (Promega 'Riboclone Eco RI Adaptor Ligation System') and the cDNA was ligated from λgt11 arms (Promega) and packaged using Stratagene 'Gigapack II Plus' packaging extracts. The phage were plated and screened following methods described by Maniatis et al. [12]. Approximately 240 000 plaques were screened using a radiolabelled PCR product as the probe. Plaque lifts were done in duplicate onto nitrocellulose filters. Filters were hybridized in 50% formamide, $5 \times$ Denhardt's solution, $5 \times$ SSPE ($1 \times SSPE = 0.15 \text{ M NaCl}, 0.01 \text{ M sodium}$ phosphate pH 7.4, 0.001 M EDTA), 0.1% SDS and 100 µg/ml denatured salmon sperm DNA at 42 °C for 18 h. The final wash was with $0.1 \times$ SSC $(1 \times SSC = 0.15 \text{ M NaCl}, 0.015 \text{ M sodium})$ citrate), 0.1% SDS at 45 °C for one set of filters and 60 °C for the second set.

A second cDNA library was constructed in the same manner, except that poly (A)⁺ RNA from carrot cell culture was used and first-strand synthesis was primed with an oligonucleotide corresponding to a tryptic peptide from purified carrot HSDH [24] (HSDH97: 5'-GCT/CTCA/GTAA/GAAA/GTAA/GTGNGTA/GTA-3').

DNA and RNA blot analysis

Genomic DNA was isolated from carrot leaves [10], digested with restriction enzymes according to the manufacturer's instructions, and $10 \mu g$ from each digestion was separated by electrophoresis on a 0.7% agarose gel and transferred to nitrocellulose by capillary action [22]. The blot was probed with the full-length AK-HSDH clone. Hybridization was in 25% formamide, $5 \times$ SSPE, $5 \times$ Denhardt's solution, 0.5% SDS and $100 \mu g/$ ml denatured salmon sperm DNA at 42 °C for 23 hr. The final wash was in $2 \times$ SSC, 0.1% SDS at 42 °C.

Total RNA was isolated from carrot cell suspension cultures at different times after inoculation and from 2-month old leaves and roots. Total RNA ($10 \mu g$) from each sample was separated on an agarose-formaldehyde-formamide gel and transferred to a nitrocellulose membrane. Gels were stained with ethidium bromide prior to transfer to confirm that equal quantities of RNA were loaded in each lane and that transfer to nitrocellulose was complete. The blot was hybridized with a ³²P-labelled 300 bp *Eco* RI-*Hinc* II fragment from the 5' end of the clone. Hybridization was in 50% formamide, $5 \times SSPE$, $5 \times Denhardt's$ solution, 0.5% SDS and $100 \mu g/ml$ denatured salmon sperm DNA at 42 °C for 18 h and washed as described above at 42 °C.

PCR amplification of a carrot cDNA encoding AK-HSDH

Protein with HSDH activity was purified from carrot cell suspension cultures [13, 24]. The purified protein was subjected to proteolysis, the polypeptides were separated by HPLC, and the amino acid sequences of four polypeptides were determined [24]. The locations of these polypeptides are presented in Fig. 1. Based on the amino acid sequence of two of these polypeptides, the following degenerate oligonucleotides were synthesized:

Oligo HSDH 76: 5'-TAT/C CAA/G GAA/G GCN TGG GAA/G ATG-3'

(Peptide 76: NH₂-Y Q E A W E M) Oligo HSDH 97: 5'-GC T/CTC A/GTA A/GAA A/GTA A/GTG NGT A/GTA-3' (Peptide 97: A E Y F Y H T Y-NH₂).

These oligonucleotides were used as primers for PCR amplification of carrot cDNA. Both double-stranded cDNA made from carrot root poly(A)⁺ RNA and first-strand cDNA made from carrot cell culture poly(A)⁺ RNA served as templates. When the amplification was carried out at an annealing temperature of 52 °C, using either template, we obtained predominantly a 1100 bp product (data not shown). The 579 bp *Eco* RI-*Hin*d III fragment from the PCR product was subcloned into vectors M13mp18 and mp19 and the DNA sequence was determined.

																				60
															CGC					
S	L	S	s	A	Ι	S	P	s	S	Y	A	A	I	A	A	A	Y	s	A	
61																				120
	GAC'	TCC	CAT	TTT	CAA	CAA	GAA	AAA	GAC	GGC	GGC	GGT'	rcT(CTC	TCC	CCT	CTC	тст	TTT	
R	T	P	I	F	N	K	K	K	T	A	A	V	Ļ	S	P	L	s	L	F	
121	TICA.	cm.c	TOO	ama	നഗന	omo.	~ 3 3	***	3.00	mam	y William	TOTAL	~~3	maa	ccc	220	***	~~ >	CIDC.	180
H						S									GGG G	AAG R	AAA K	GGA E	S	
••	*	_	•	_		-	•	-	~	-	•	_		-	•			-	~	
181																				240
															TTC					
S	s	K	F	Y	Ι	A	A	s	V	T	т	A	V	P	s	L	D	D	s	
241																				300
	TGA	GAA	сст	ሞሮል	ССТ	TCC	CAG	ccc	TGC	ጥልጥ	стс	STO	ידאיו	TCA	TAA	ארת	TGG	AGG	CAC	300
V						P						s	I	H		F	G	G	T	
															_	_	_		_	
301																				360
															TGT				TTC	
С	V	G	S	S	Е	R	1	ĸ	N	٧	A	Е	Ι	٧	٧	E	D	D	5	
361																				420
	AAG.	AAA	GCT	AGT	TGT	AGT	CTC	TGC	AAT	GTC.	AAA	GGT	CAC	AGA	TAT	GAT	GTA	TGA	TCT	
E						V				S		v	\mathbf{T}		M	M	Y	D	L	
421														a		ma.	mam		~~~	480
									TTC S			ATC' S		GCI L	CGA		TGT V	TAT M	GGA E	
I	I	K	A	Q	3	R	D	υ	3	1	ь	3	A	_	D	A	٧	м	£	
481																				540
	GCA	CAA	GTT	AAC	AGC	ATT	TGA	TCT	CCT	TGA'	TGA.	AGA'	TGA	CCI	'TGC	TAG	ATT	TTT	AAC	
K	H	K	L	T	A	F	D	L	L	D	E	D	D	L	A	R	F	L	T	
541	- cm			~~~	mam			m				a ami		ma.		3 m 3	a		maa	600
	SCIT	QCA.		TGA D		TAT I	TAC T			AGC A	AAT'	E L	reg R	TGC A	TAA: I	ATA Y	I	AGC A	G	
R	יו	Q	п	D	٧	+	1	L	М	A	м	ъ	К	A	+	1	1	A	G	
601																				660
	TGC	CAC	CGA	ATC	TTT	TTC	GGA	TTT	TGT	TGT	GGG.	ACA'	TGG.	AGA	GCT	ATG	GTC	AGC	TCA	
H	Α	T	E	S	F	S	D	F	V	V	G	H	G	E	L	W	s	A	Q	
661																				
	amm.	ama	3 mm	man	3 3 m			~ · · ·	maa	~~~	max.	amai	7122	mmz		~~3	~~			720
GCT															GAT M					720
	GTT L			TGT V			AAA K		TGG G		TGA D			TTG W	GAT M	GGA D	CAC T	ACG R	AGA D	720
GCT																				720 780
GCT L 721 TGT	L FCT'	S TGT	F TGT	V AAA	I TCC	R TGC	K TGG	N ATC	G TAA	G TCA	D AGT	C CGA	n rcc	W TGA		D	T	R	D	
GCT L 721 TGT	L FCT'	S TGT	F TGT	V AAA	I TCC	R	K TGG	N ATC	G TAA	G TCA	D AGT	C CGA	n rcc	W TGA	M	D	T	R	D	
GCT L 721 TGT	L FCT'	S TGT	F TGT	V AAA	I TCC	R TGC	K TGG	N ATC	G TAA	G TCA	D AGT	C CGA	n rcc	W TGA	M TTA	D TTT	T GGA	R ATC	D TGA	
GCT L 721 TGT	L FCT'	S TGT	F TGT	V AAA	I TCC	R TGC	K TGG	N ATC	G TAA	G TCA	D AGT	C CGA	n rcc	W TGA	M TTA	D TTT	T GGA	R ATC	D TGA	
GCT L 721 TGT V 781	L PCT' L	s TGT' V	F TGT V	V AAA N	I TCC P	R TGC A	K TGG G	N ATC S CAG	G TAA N	G TCA Q CCA(D AGT V	C CGA' D	N TCC P	W TGA D	M TTA Y TGT:	D TTT L IGC	T GGA E	R ATC S	D TGA E TTT	780
GCT L 721 TGT V 781	L PCT' L	s TGT' V	F TGT V	V AAA N	I TCC P	R TGC A	K TGG G	N ATC S CAG	G TAA N	G TCA Q	D AGT V	C CGA' D	N P	W TGA D	M TTA Y	D TTT L	T GGA E	R ATC S	D TGA E	780
GCT L 721 TGT V 781 GAAC K	L FCT' L	S TGT V	F TGT V TGA	V AAA N	I TCC P	R TGC' A GTT(K TGG G	N ATC S CAG	G TAA N	G TCA Q CCA(D AGT V	C CGA' D	N TCC P	W TGA D	M TTA Y TGT:	D TTT L IGC	T GGA E	R ATC S	D TGA E TTT	780 840
GCT L 721 TGT V 781 GAAG K	L ICT' L SAG! R	S TGT'' V ACT''	F V V rga E	V AAA N SAA K	TCC P ATGG	TGC A STTC F	K TGG. G CTCC	N ATC S CAG	G TAA N CAA	G TCA Q CCAC Q	D AGT V STG1 C	C CGA' D CCAC	N P P GAC! T	W TGA D	M TTA Y TGT:	TTT L IGC A	GGA E GAC	ATC S AGG G	TGA E TTT F	780
GCT L 721 TGT V 781 GAAC K	L ICT' L SAGA R	TGT V ACT L	F TGT V TGAC E	V AAA N GAA K	TCC P ATGG W	TGC A STTC F	K TGG G CTCC S	ATC S CAG S	G TAA N CAA N	G TCA: Q CCA: Q	AGTO V STGT C	C CGA' D CCAC Q	N P GACI T	W TGA D	M TTA Y TGT:	TTT L IGC A	GGA E GAC	ATC S AGG G	TGA E TTT F	780 840
GCT L 721 TGT V 781 GAAG K	L ICT' L SAG! R	S TGT'' V ACT''	F V V rga E	V AAA N SAA K	TCC P ATGG W	TGC A STTC F	K TGG. G CTCC	N ATC S CAG	G TAA N CAA	G TCA Q CCAC Q	D AGT V STG1 C	C CGA' D CCAC	N P P GAC! T	W TGA D AAT I	M TTA Y TGT V	TTT L IGC A	GGA E GAC: T	ATC S AGG G	TGA E TTT F	780 840
GCT L 721 TGT V 781 GAAC K	L ICT' L SAGA R	TGT V ACT L	F TGT V TGAC E	V AAA N GAA K	TCC P ATGG W	TGC A STTC F	K TGG G CTCC S	ATC S CAG S	G TAA N CAA N	G TCA: Q CCA: Q	AGTO V STGT C	C CGA' D CCAC Q	N P GACI T	W TGA D AAT I	M TTA Y TGT V	TTT L IGC A	GGA E GAC: T	ATC S AGG G	TGA E TTT F	780 840
GCT L 721 TGT V 781 GAAC K 841 I 901 TGCC	L TCT' L BAGG R	S TGT V ACTT L FAGG	F TGT: V TGAC E CAC	V AAAA N GAAA K GGCC' P	TCC P ATGC W TCA	R TGC A GTT(F AAAA: N	K TGG G CTCC S	N ATC S CAG S ACC	G TAA N CAA N TAC	G TCAI Q CCAC Q T T TIGGT	D AGTOV C C C C C C C C C C C C C C C C C C C	C CGA D PCAC Q Q	N TCC P SACA T AAAGA R	W TGA D AAAT I AGA D	M TGTTA V CGG	D TTTT L IGCO A AAG	T GGA E GAC T TGA	R ATC S AGG G CTT F	TGA E TTT F TTC S	780 840 900
GCT L 721 TGT V 781 GAAC K 841 I 901 TGCC	L TCT' L BAGG R	S TGT V ACTT L FAGG	F TGT: V TGAC E CAC	V AAAA N GAAA K GGCC' P	TCC P ATGC W TCA	R TGC A GTT(F AAAA: N	K TGG G CTCC S	N ATC S CAG S ACC	G TAA N CAA N TAC	G TCAI Q CCAC Q T T TIGGT	D AGTOV C C C C C C C C C C C C C C C C C C C	C CGA' D PCAC Q Q	N TCC P SACA T AAAGA R	W TGA D AAAT I AGA D	M TTA Y TGT V CGG	D TTTT L IGCO A AAG	T GGA E GAC T TGA	R ATC S AGG G CTT F	TGA E TTT F TTC S	780 840 900
GCT L 721 TGT V 781 GAAC K 841 TATA I 901 TGC A	L TCT' L BAGG R	S TGT V ACTT L FAGG	F TGT: V TGAC E CAC	V AAAA N GAAA K GGCC' P	TCC P ATGC W TCA	R TGC A GTT(F AAAA: N	K TGG G CTCC S	N ATC S CAG S ACC	G TAA N CAA N TAC	G TCAI Q CCAC Q T T TIGGT	D AGTOV C C C C C C C C C C C C C C C C C C C	C CGA' D PCAC Q Q	N TCC P SACA T AAAGA R	W TGA D AAAT I AGA D	M TGTTA V CGG	D TTTT L IGCO A AAG	T GGA E GAC T TGA	R ATC S AGG G CTT F	TGA E TTT F TTC S	780 840 900 960
GCT L 721 TGT V 781 GAAC K 841 TATA I 901 GGAAC A 961	L TCT' L AAGC' A	S TGT V ACTT L FAGG	F TGT. V TGAC T	V AAA N GGCC P GGGG	TCC P ATGC W TCA Q TGC	R TGC' A GTT(F AAAA: N	K TGG G CTCC S FATT I	N ATC S CAG S ACC P	G TAA N CAA N TAC T GGC'	G TCA Q CCAC Q T T TGGT	D AGT V STGT C C TTTC L CCAA	C CGA' D CCAC Q Q SAAA K	N P GAC! T AAAG! R	W TGA D AAAT I AGA D	M TTA Y TGT V CGGG G	D TTTT L TGC: A AAG S	GGAC EGAC TGA D	R ATC S AGG G CTT F	D TGA E TTT F TTC S TAA	780 840 900
GCT L 721 TGT V 781 GAAC K 841 TATT I 901 TGC A 961	L ICT' L SAGA R AGCT A	S TGT V ACTT L FAGG	F TGT. V TGAC E CACC T	V AAA N GGCC P GGG G	TCCC P ATGC W TCAI Q TGC: A	R TGCTA AAAAT N TTTTT	K TGGG G CTCC S FATTI	N ATC S CAG S ACC P	G TAA N CAAN N TAC	G TCA. Q CCAC Q TAGG	D AGT V FTTC	C CGA' D PCAC Q V	N TCC P SACA T CACC T	W TGA D AAAT I AGA D GAT I	M TTA Y TGT V CGG G TTGG W	D TTTT L FGCO A AAGG S GACGT T	T GGA E GAC T TGA D	R ATC S AGG G CTT F	D TGA E TTT F TTC S TTAA N	780 840 900 960
GCT L 721 TGT V 781 GAAC K 841 TATT I 901 TGC A 961	L ICT' L SAGA R AGCT A	S TGT V ACTT L FAGG	F TGT. V TGAC E CACC T	V AAA N GGCC P GGG G	TCCC P ATGC W TCAI Q TGC: A	R TGCTA AAAAT N TTTTT	K TGGG G CTCC S FATTI	N ATC S CAG S ACC P	G TAA N CAAN N TAC	G TCA. Q CCAC Q TAGG	D AGT V FTTC L C C C C C C C C C C C C C C C C C	C CGA' D PCAC Q V	N TCC P SACA T CACC T	W TGA D AAT I AGA D GAT I GGT V	M TTA Y TGT V CGGG G	D TTTT L TGC: A AAAG T AAAA	T GGAC T TGA D GAC T	R ATC S AGG G CTT F TGT V ATT L	D TGA E TTT F TTC S TTAA N	780 840 900 960
GCT L 7211 TGT V 781 GAAC K 841 I TTATI I 901 TGGC A	L FCT'L R AAGC'A A	S TGT V ACTT L FAGG	F TGT. V TGAC E CACC T	V AAA N GGCC P GGG G	TCCC P ATGC W TCAI Q TGC: A	R TGCTA AAAAT N TTTTT	K TGGG G CTCC S FATTI	N ATC S CAG S ACC P	G TAA N CAAN N TAC	G TCA. Q CCAC Q TAGG	D AGT V FTTC L C C C C C C C C C C C C C C C C C	C CGA' D PCAC Q V	N TCC P SACA T CACC T	W TGA D AAT I AGA D GAT I GGT V	M TTA Y TGT V CGG G TTGG W ATT L	D TTTT L TGC: A AAAG T AAAA	T GGAC T TGA D GAC T	R ATC S AGG G CTT F TGT V ATT L	D TGA E TTT F TTC S TAA N GTC S	780 840 900 960
GCT L 721 TGT V 781 GAAG K 841 TATT I 901 TGGG G 102:	L FIGTI L SAGA R AGCT A L FIGTI V	S TGTT V ACTT L FAGG S FATAT I	F TGT: V TGAC E CAC T TAAT M	V AAAA N GCCC P GGGG G	TCC P ATGO W TCAA Q TGCA A AGA	R TGC' A AAAT N TTTTT L	K TGG. G CTCC S FATA I L TCGA R	N ATC S CAG S ACC P AAAG R	G TAA N CAAC N TAC: T GGCC A AGT	G TCA. Q CCCAC Q AAACT T TGGT G	D AGTGT C C TTTC L CCAA Q CGAC E	C CGA' D CCAC Q SAAA K V	N FICC' P FAAGA T CACC T COMMENT V	W TGA D AAT I AGA D GAT I GGT V	M TTTA Y TGT V CGG, G TTGG W	D TTTT L IGCC A AAGC T AAAA K 76	T GGA E GAC T TGA D GAC T T	R ATC S AGG G CTT F TGT V ATT L L	D TGA E TTT F TTC S TTAA N GTC S S	780 840 900 960
GCT L 721 TGT V 781 GAAC K 841 TTTGC A 961 TTGG G	L FIGTI L SAGA R AGCT A L FIGTI V	S TGT V ACTT L FAGG S FATAT Y	F TGT: V TGAC: E CAC: T TAAT: M	V AAAA N GAA. K GCC P GGG G G CTGC	TCC P ATGC W TCAA Q TGCA A AGA D	R TGC' A AAAT N TTTTT L TCC' P	K TGG. G TATA I ATTA L TCGA	N ATC S CAG S ACC P AAAG R AAAA	G TAA N CAAN N TAC:	G TCA. Q CCAC Q AAAC T TGGT G S TTGGG	D AGTGT V STGT C C L CCAA Q CGAC E	C CGA' D CCAC Q Q SAAA K V V	N FICC' P FAAGA R R CACC T FIGTO V	W TGA D AAT I AGA D GAT V GTT	M TTTA Y TGTT V CGGG G TTGG W ATT L Pep	D TTTT L IGCC A AAAG S GAC' T AAAA K 76	T GGA E GAC T TGA D GAC T T CCG	R ATC S AGG G CTT F TGT V ATT L L	D TGA E TTC S TTAA N GTC S TAT	780 840 900 960
GCT L 721 TGT V 781 GAAC K 841 TTATI I 901 TGGC G	L TCT' L SAGA R AGCT A L TGTI V	S TGT V ACTT L FAGG S FATAT Y AGA E	F TGT V TGAC E CACC T TAAAT M	V AAAA N GGCC P GGG G G CTGC W	TCC P ATGO W TCAA Q AGA D GGAA	R TGC A STTC F AAAA N TTTT L GATC M GATC	K TGG G TTCG S TATTI L TCGI R	N ATC S CAG S ACC P AAAG R AAAA K	G TAA N CAAN N TAC. T GGC. A AGT. V	G TCAL Q CCAC Q TGGT G TTGGT G TTGGGC G	D AGTOV FTTO C C FTTO L FCAA Q FGAO E GGCC A	C CGA' D CCGA' C C CCGA' C C C C C C C C C C C C C C C C C C C	N TCC' P GACI T AAAGI R CACCT V FGTC V	W TGA D AAT I AGA D GAT V GTT L	M TTA Y TGT V CGGG G TTG W ATT L Pep	D TTTT L TGCCA AAAA K 76 TCCC	T GGA E GAC T TGA D GAC T T CCG R	R ATC S AGG G CTT F TGT V ATT L L	D TGA E TTC S TTAA N GTC S TAT	780 840 900 960
GCT L 721 TGT V 781 GAAC K 841 TTTGC A 961 TTGG G	L TCT' L SAGA R AGCT A L TGTI V	S TGT V ACTT L FAGG S FATAT Y AGA E	F TGT V TGAC E CACC T TAAAT M	V AAAA N GGCC P GGG G G CTGC W	TCC P ATGO W TCAA Q AGA D GGAA	R TGC' A AAAT N TTTTT L TCC' P	K TGG G TTCG S TATTI L TCGI R	N ATC S CAG S ACC P AAAG R AAAA K	G TAA N CAAN N TAC. T GGC. A AGT. V	G TCAL Q CCAC Q TGGT G TTGGT G TTGGGC G	D AGTOV FTTO C C FTTO L FCAA Q FGAO E GGCC A	C CGA' D CCAC Q Q SAAA K V V	N TCC' P GACI T AAAGI R CACCT V FGTC V	W TGA D AAT I AGA D GAT V GTT L	M TTA Y TGT V CGGG G TTG W ATT L Pep	D TTTT L IGCC A AAAG S GAC' T AAAA K 76	T GGA E GAC T TGA D GAC T T CCG	R ATC S AGG G CTT F TGT V ATT L L	D TGA E TTC S TTAA N GTC S TAT	780 840 900 960
GCT L 721 TGT V 781 GAAC K 841 TTAT I 901 TGGG G G 102: TTTAT Y Y	L ICT' L SAGA R AGCT A CGGTI V L CGGTI Q Q L	S TGT V ACTT L FAGG S FATAT L ATAT	F TGT. V FGAC E CACC T AATC M	V AAAA N SAAA K GCC P GGG G G G CTG W	TCCC P ATGC W TCAA Q AGA D GGA E E	R TGC' A AAAA' N TTTTI L TCC' P GATG M	K TGG G CTCC S FATTI I ATTT L CGG S S S	N ATC S CAG S ACC P AAAG R AAAA K ATA Y	G TAAA N TAC: T T GGGC' V TTTT' F	G TCAL Q CCAC Q AAAC T TIGGT S TAGC G G G G G G	D AGT	C CGA'D CCAC Q C CAC X X X X X X X X X X X X X X X X X	N ICCC P AAGA T AAGA R CACC T V V V V V	W TGA D AAT I AGA T V GTT L L	M TTA Y TGT' V CGGG G TTG W ATT; L Pep ACA' H	D TTTT L IGCO A AAAC S GAC T AAAA K 76 P P	GGACAT T CCCGRR	R ATC S AGG G CTT F TGT V ATT L TAC T	D TTGA E TTTT F TTC S TTAA N GTC S TAT I	780 840 900 960
GCT L 721 TGT V 781 GAAC K 841 TATT I 901 TGGG G 102: TTTA' Y 108: CAA	L ICT' L SAGA R AGCT A CGCT A CGCT L CCAL Q Q	S TGT V ACTT L FAGO S FATA Y AGAI E E	F TGT. V TGAC T TAAT M TAG S	V AAAA N SAAA K GCCC P GGG G CTGC W W	TCCC P ATGC W TCAA Q TGC A D GGAA E E ATA	R TGC A GTTC F AAAA N FTTT L GATC M M TGA	K TGG G CTCC S FATTI I CGG R S CATC	N ATC S CAGG P AAAG R AAAA K ATA Y Y TCC	G TAAA N CAAA N TACA T T GGGC A V TTTT F F	G TCAL Q CCCAC Q AACT T TGGG G G G G G TTGGG	D AGTOV STGT C C PTTCAL Q PGACE A A AATA	C CGA'D CCAC Q CAGTO X K AGTO X N N N AAAGA	N ICCC P SACA T AAAGA R CACC T V V V V V V V V V V V V V V V V V V	W TGA	M TTA Y TGTT V CGG. G TTGG W ATT. Pep ACA H H	D TTTT L TGCCA AAAA K 76 P P CAA	GAC. TTGA' D GAC. TTGA'	R ATC S AGG G CTT F TGT V ATT L TAC T	D TGA E TTT F TTC S TAA N GTC S S TAT I	780 840 900 960 1020
GCT L 721 TGT V 781 GAAC K 841 TATT I 901 TGGG G 102: TTTA' Y 108: CAA	L ICT' L SAGA R AGCT A CGCT A CGCT L CCAL Q Q	S TGT V ACTT L FAGO S FATA Y AGAI E E	F TGT. V TGAC T TAAT M TAG S	V AAAA N SAAA K GCCC P GGG G CTGC W W	TCCC P ATGC W TCAA Q TGC A D GGAA E E ATA	R TGC A GTTC F AAAA N FTTT L GATC M M TGA	K TGG G CTCC S FATTI I CGG R S CATC	N ATC S CAGG P AAAG R AAAA K ATA Y Y TCC	G TAAA N CAAA N TACA T T GGGC A V TTTT F F	G TCAL Q CCCAC Q AACT T TGGG G G G G G TTGGG	D AGTOV STGT C C PTTCAL Q PGACE A A AATA	C CGA'D CCAC Q CAGTO X K AGTO X N N N AAAGA	N ICCC P SACA T AAAGA R CACC T V V V V V V V V V V V V V V V V V V	W TGA	M TTA Y TGT' V CGGG G TTG W ATT; L Pep ACA' H	D TTTT L TGCCA AAAA K 76 P P CAA	GAC. TTGA' D GAC. TTGA'	R ATC S AGG G CTT F TGT V ATT L TAC T	D TGA E TTT F TTC S TAA N GTC S S TAT I	780 840 900 960 1020
GCT L 721 TGT V 781 GAAC K 841 TTAT I 901 TGGG G G 102: TTA: Y Y 108: CAAC N	L FCT. L SAGA R AGCTA A CGCTA V L CCAT P	S TGT V ACTT L FAGO S FATAT Y AGAIN E E	F TGT. V TGAC T TAAT M TAG S	V AAAA N SAAA K GCCC P GGG G CTGC W W	TCCC P ATGC W TCAA Q TGC A D GGAA E E ATA	R TGC A GTTC F AAAA N FTTT L GATC M M TGA	K TGG G CTCC S FATTI I CGG R S CATC	N ATC S CAGG P AAAG R AAAA K ATA Y Y TCC	G TAAA N CAAA N TACA T T GGGC A V TTTT F F	G TCAL Q CCCAC Q AACT T TGGG G G G G G TTGGG	D AGTOV STGT C C PTTCAL Q PGACE A A AATA	C CGA'D CCAC Q CAGTO X K AGTO X N N N AAAGA	N ICCC P SACA T AAAGA R CACC T V V V V V V V V V V V V V V V V V V	W TGA	M TTA Y TGTT V CGG. G TTGG W ATT. Pep ACA H H	D TTTT L TGCCA AAAA K 76 P P CAA	GAC. TTGA' D GAC. TTGA'	R ATC S AGG G CTT F TGT V ATT L TAC T	D TGA E TTT F TTC S TAA N GTC S S TAT I	780 840 900 960 1020
GCT L 721 TGT V 781 GAAC K 841 FTATI I 901 FTGC A 102: TTTA Y Y 108: CAAN N 114:	L FCT' L SAGAR R AGCT A L FCCAI Q Q L FCCC P	S TGT V ACTT L FAGG S FATAT Y AGA E E	F TGT. V TGAC T TAAAT M TAG S	V AAAA N GCC P GGG G GCG W W GCCG R	TCCP ATGCW TCAA Q TGCA AGA D GGAA E E ATA	R TGC' A AAAA' N TTTTI L GATC M M TTGAC	K TGGG G TATT I TCGG R GTCG S S CATT I	N ATC S CAGE S ACC P AAAG K ATA Y Y ICC	G CAAN N TACL T T F F AAT I	G TCAL Q CCAC Q AAC T TGGG G G TTGGC G G TTGGT V	D AGT V STGT C C TTTC L C C C C C C C C C C C C C C	C CGA'D CCAC Q Q GAAAAGT V V FAAST N N N AAAGA R	N FICC' P AAGA T CACC T FGTO V V AAAA	W TGA TGA T T T T T T T T T T T T T T T T	M Y TGT V CGG G TTG W ATT L Pep ACA H H	D TTTT L IGCCA AAG S GAC T AAAA K 76 P CAA	GGACT T GACT T CCGGRR	ATC S AGG G CTT V ATT L TAC T ATC	TGA F TTT F TTC S TAA N GTC S S	780 840 900 960 1020
GCT L 721 TGT V 781 GAAC K 841 ITATI 901 TGCC A 961 TGG' Y Y 108: CAA' N	L FICT' L AGCT A FICAL FIC	S TGT V ACTT L FAGO S FATA Y AGAI E E FGTO V	FIGT. V FIGACIE E CACCIT T AATTO S AGCIA A AATTO M AAATTO M AAATTO	V AAAA N GAAL K GCC P GGG G GGR W W GCG R GAT	TCCCP ATGCCA W TCAA Q TGCCA A AGA TGCA A AGA TGCA A A A A A A A A A A A A A A A A A A	R TGCT A AAAAT L TCCT P GATC M M TGAC	K TGG G CTCC S FATT I FCG R GTC S CATT I	N ATC S CAG'S ACC P AAAG R AAAA K ATA'Y Y TCCC P ATC	G TAA N CAAN N TACL T GGC A V TTTT F F AAT I TGT.	G TCA. Q CCCAC Q TTGGTT G G G TTGGC G G TGTT V	D AGT V STGT C C TTTCA A A A A A A T I	C CGA' D CCAC Q C C C C C C C C C C C C C C C C	N FGTO FGTO V V AAAA	W TGA D AAT I AGA D GTT L AGA AGA	M TTA Y TGTT V CGG. G TTGG W ATT. Pep ACA H H	D TTTT L TGCCA AAAG S GAC T CCC P P CAA N	GGACT TGAT D CCG R R CCT.	R ATC S AGG G CTT V ATT L ATC S ATT	TGA F TTT F TTC S TAA N GTC S S	780 840 900 960 1020
GCT L 7211 TGT V 781 GAAC K 841 TTTTT I 9011 TGGC A 961 TGGC Y Y 108: CAAC N	L FGTA CGCT A CGCT A CGCT CGC	S TGT V ACTT L FAGO S FATA Y AGAI E E FGTO V	FIGT. V FIGACIE E CACCIT T AATTO S AGCIA A AATTO M AAATTO M AAATTO	V AAAA N GAAL K GCC P GGG G GGR W W GCG R GAT	TCCCP ATGCCA W TCAA Q TGCCA A AGA TGCA A AGA TGCA A A A A A A A A A A A A A A A A A A	R TGCT A AAAAT L TCCT P GATC M M TGAC	K TGG G CTCC S FATT I FCG R GTC S CATT I	N ATC S CAG'S ACC P AAAG R AAAA K ATA'Y Y TCCC P ATC	G TAA N CAAN N TACL T GGC A V TTTT F F AAT I TGT.	G TCA. Q CCCAC Q TTGGTT G G G TTGGC G G TGTT V	D AGT V STGT C C TTTCA A A A A A A T I	C CGA' D CCAC Q C C C C C C C C C C C C C C C C	N FGTO FGTO V V AAAA	W TGA D AAT I AGA D GTT L AGA AGA	M TTA Y TGT' V CGGG G TTG W ATT L Pep ACA' H H TGGG	D TTTT L TGCCA AAAG S GAC T CCC P P CAA N	GGACT TGAT D CCG R R CCT.	R ATC S AGG G CTT V ATT L ATC S ATT	D TGA E TTT F TTC S TAA N GTC S TAT I TGC A GGA	780 840 900 960 1020 1080
GCT L 721 TGT V 781 GAAC K 841 TTGTC A 961 TTGCC A 102: TTA Y 108: CAA: N 114:	L ICT' L R AAGCT A PGTT V L CGCT P L CGGGG	S TGTT V ACTT L FAGO S FATA I ATA Y AGA E E T T AAACA T	FIGT V FGACE ECACC T T AAATC M FAG AAATC M AAATC M AAATC M AAATC M	V AAAA N GAAA K GCCC P GGGG G G G G G G G G G G G G G G	TCCC P ATGC W TCAA Q TGCC A D GGAA AGA: Y ATGC	R TGCT A AAAA N TTTT L TCCT P GATC M M TGAC R	K IGG G CTCC S FATTI I CTCG R CATTI AGA E	N ATC S CAG P AAG R ATA Y Y TCC P ATC S	G TAA N CAAN N TAC. T GGC A A GT F F F T GT. V	G TCA. Q CCAC Q TTGGTT G G G G G TGGTT V	D AGT V V STGT C C C C C C C C C C C C C C C C C C	C CGAC D CCGAC Q C CGGC N N N N AAAGT T T	N FGAAAGA T CACC T FGTC V V AAAAT E	W TGA D AAT I AGA D GTT L AGA D AGA D	M TTA Y TGT V CGG G TTG W ATT L Pep ACA H H TGG G	D TTT L IGCO A AAG S GAC T AAAA 76 P CAA N GTT L	GGACT TGACT	ATC S ATT L ATC S ATT L	TGA F TTT F TTC S TAA N GTC S S TAT I TGC A E GGA	780 840 900 960 1020
GCT L 7211 TGT V 781 GAAC K 8411 TTGT I 9011 TGGC A 9611 TGGC A 102: TTTAT Y Y 108: CAA N 1114: CAA 114: CAA 11	L ICT' L SAGA R AGCT A CGCT V L CCCT P L GGGG G L CCCT CCAT	S TGT V ACTT L FAGG S FATA I ATA Y AGA E E TGT V AACLT TGT TGT TGT TGT TGT TGT TGT TGT TGT T	FIGT. V FGACE CACCO T CACCO T AATCO M FAGC S AGC A A A A A A A A A A A A A A A A A A	V AAAA N GAAA K GCCC P GGG G G G G G G G G G G G G G G	TCCCP ATGC W TCAA Q TGC A AGA AGA AGA Y ATGC A ATT	R TGCT F AAAAT N TTTT L TCCT P GATCA M M TGAC R TGCT R TGCT R	K TGG G CTCC S FATTI I ATTI L FCG S S CATTI I AGA E	N ATC S CAGE S ACC P AAAA K ATA Y Y FCC P TATC	G CAA'N TAC'N TAC'N TTTT'F F AAT'I TGT.	G TCAC Q CCAC Q AAAC T TGGT G G G G G TTAG X AGGG G G TTAA	D AGTY V FITTO C FTTO Q FIGAC E A A A A T I C G A A A T I C G A A A A T I C C A A A A A T I C C A A A A A A A A B B C B C C A A A A	C CGA'D CCGA'C CGA'C CGA'C CGGCC A CGCC	N FCC P T T AAGA R CACO T V V V V AAAA: E GCT C GCT C O O O O O O O O O O O O O O O O O O	W TGA D AAT I AGA D GTT L AGA D CAT	M TTA Y TGT' V CGGG G TTG W ATT L Pep ACA' H H TGGG	D TTT L IGC A AAG S GAC T C C C C C C C C C C C C C C C C C C	T GGACT T CCGGR R C C L AAAA K	R ATC S AGG G CTT F TGT V ATT L ATC T ATC S ATT AGG	D TGA TTC S TAA N GTC S S TAT I TGC A AAC	780 840 900 960 1020 1080

126																				
																				1261
TGG	AAT	GGC'	TGG'	rgt'	TCC	TGG'	TAC	AGC	TAC	TGC.	AAT'	ltt	TGG	TGC	TGT	CAA	GGA	TGT	GGG	
G	M	Α	G	V	P	G	\mathbf{T}	A	Т	A	I	F			V		D	V	G	
G	M	A	G	٧	P	G	т	Α	T	Α	Ι	F	[S]	Α	V	(K)				
	_																			
132		T-C-M				3 m 🚓														1380
																CTT				
A	N	V	I	M	Ι	S	Q	A	s	S	E	н	S	1	С	F	A	V	P	
120																				
138	_	T		~																1440
																TCA				
E	S	E	٧	K	А	٧	Α	K	A	L	E	Α	R	F	R	Q	A	L	D	
1 4 4	-																			
144		T00	TOM	rmo.		-a						-m-		~~ ~						1500
TGC																				
A	G	к	ь	S	Q	٧	A	N	N	Р	N	С	s	I	L	A	\mathbf{T}	V	G	
150	1																			1560
CCA	AAA	SAT	GGC2	\AG'	rac'	rcc'	rgg.	CGT	GAG'	rgc:	PAC	ACT	TTT	CAA	TGC	GCT'	TGC	AAA	GGC	
Q	K	M	A	S	Т	P	G	v	S	Α	T	L	F	N	Α	L	Α	K	Α	
156	1																			1620
CAA'	TAT	AAA	CGT'	rcg:	rgc:	rat.	AGC	CCA	GGG	CTG:	PAC	AGA	GTA	TAA	TAT	CAC'	TGT.	AGT'	TCT	
N	I	N	v	R	A	I	Α	Q	G	С	T	E	Y	N	I	T	v	V	L	
162	1																			1680
CAG	TCG	AGA	AGA	rTG	rgr	GAG	GGC	TTT	GAA	AGC:	rgre	CCA	TTC	AAG	ATT	TTA'	TCT	GTC	GAG	
S			-						K							Y	L	S	R	
_			_	-	-			_			-		_		_	_				
168	1																			1740
AAC		ΔАТ	AGC	CTC	agg:	יידאיו	тст	CGG	ACC	rggi	4 TT	AΑT	CGG	AGC	TAC	ттт	АСТ	TGA	CCA	_,,,
	T		A						P				G		T		L		Q	
•	-	-	••	•	•	-	•	•	-	_	_	-	•	••	-	_		_	×	
174	1																			1800
		262	י ביים	ccc	A C C	ል አጥሪ	രസ	ממח	CCN:	ימממ	יייייי	гаа	λλπ	ጥርል	արդ	'GCG'	ጥረጥ	ጥልጥ	acc	1000
L			Q						E						L	R	v	M	G	
-	К	U	×	Α	Ω	-	п	1	-	14	5	r	_				٧	11	G	
180	1																			1860
		~~~	a mod	720		2 2 m	com	mam	CNC	003			3 3 171	003	mmm	330	ma c	a mod	C 3 C	1000
TAT																				
I	T	G	s	ĸ	T	M	ь	ъ	s	E	T	G	Ι	D	L	S	R	W	R	
106																				1020
186		CC3	***		~ ~ ~ ~		~~~	220	300	TICC/	20m			N TOTAL	men	202	202	mem.	ccc	1920
AGA.			K									AGA E							R	
Ľ,	V	Q	r	L	М	G	Q	1	A	G	ь	E	V	r	٧	Q	п	٧	К	
100	-																			1000
192	_	TO 3 (	nam	n 2 m	naa-			mam	m 3 m			nma	m 2 0				ma 3	3 am		1980
TGG.																				
G				1		S	Т	V		V	ט	C	т	A	ע	S	Е	v	A	
	IA	H	F	_	F															
		Н	F	_	r															
198	1																			2040
AAG	1 TCA	ста	CCA!	rga:	CTG				GGG.							'AAA'				2040
AAG	1 TCA		CCA!	rga:												'AAA' N		GAA K	GGC A	2040
AAG S	1 TCA: H	ста	CCA!	rga:	CTG				GGG.											
AAG S 204	1 TCA H	CTA Y	CCA'	rga D	CTG W	Ļ	С	R	GGG; G	Ι	H	V	I	Т	P	N	K	K	A	2040
AAG S 204 AAA	1 TCA: H 1 TTC:	CTA Y AGG	CCA! H	rga D	CTG W TGA	L TCA	C GTA	R TTT	GGG. G	I GTT	H GAG	V AGC	I TCT	Т	P .GCG	n GCG.	K ATC	K CTA	A	
AAG S 204 AAA	1 TCA: H 1 TTC:	CTA Y AGG	CCA! H	rga D	CTG W TGA	L TCA	C GTA	R TTT	GGG; G	I GTT	H GAG	V AGC	I TCT	T CCA Q	P .GCG R	N GCG. R	K ATC S	K CTA Y	A TAC T	
AAG S 204 AAA	1 TCA: H 1 TTC:	CTA Y AGG	CCA! H	rga D	CTG W TGA	L TCA	C GTA	R TTT	GGG. G	I GTT	H GAG	V AGC	I TCT	T CCA Q	P .GCG R	n GCG.	K ATC S	K CTA	A TAC	
AAG S 204 AAA	1 TCA: H 1 TTC:	CTA Y AGG	CCA! H	rga D	CTG W TGA	L TCA	C GTA	R TTT	GGG. G	I GTT	H GAG	V AGC	I TCT	T CCA Q	P .GCG R	N GCG. R	K ATC S	K CTA Y	A TAC T	
AAG S 204 AAA N 210	1 TCA H 1 TTC S	CTA Y AGG. G	CCAT H ACCO P	rgae D CCT' L	CTG W I'GA' D	L TCA Q	C GTA Y	R TTT L	GGG. G GAA K	I GTT( L	H GAG R	V AGC A	I TCT L	T CCA Q	P .GCG R <u>Pe</u> p	N GCG R 97	K ATC S S	K CTA' Y Y	A TAC T	
AAG S 204 AAA N 210	1 TCA H 1 TTC S	CTA Y AGG. G	CCAT H ACCO P	rgae D CCT' L	CTG W I'GA' D	L TCA Q	C GTA Y	R TTT L	GGG. G GAA K	I GTT( L	H GAG R	V AGC A	I TCT L	T CCA Q	P .GCG R <u>Pe</u> p	N GCG. R	K ATC S S	K CTA' Y Y	A TAC T	2100
AAG S 204 AAA N 210 ACA H	1 TCA H TTC S S 1 CTA	CTA Y AGG. G TTT	CTA	rga D CCT L FGA	CTG W TGA' D AGC' A	L TCA Q TAC	C GTA Y TGT V	R TTT L TGT V	GGG. GAA K TGC'	I GTT( L TGG	H GAG R PCT	V AGC A	I TCT L GAT	T CCA Q CAT	P GCG R Pep	N GCG R 97	K ATC S S	K CTA' Y Y GCA	A TAC T T	2100
AAG S 204 AAA N 210 ACA H	1 TCA H TTC S S 1 CTA	CTA Y AGG. G TTT	CTA	rga D CCT L FGA	CTG W TGA' D AGC' A	L TCA Q TAC	C GTA Y TGT V	R TTT L TGT V	GGG. G GAA K TGC	I GTT( L TGG	H GAG R PCT	V AGC A	I TCT L GAT	T CCA Q CAT	GCG R Pep	N GCG. R 97	K ATC S S TTT	K CTA' Y Y GCA	A TAC T T GGG	2100
AAG S 204 AAA N 210 ACA H	1 TCA H TTC S S 1 CTA	CTA Y AGG. G TTT	CTA	rga D CCT L FGA	CTG W TGA' D AGC' A	L TCA Q TAC	C GTA Y TGT V	R TTT L TGT V	GGG. GAA K TGC'	I GTT( L TGG	H GAG R PCT	V AGC A	I TCT L GAT	T CCA Q CAT	GCG R Pep	N GCG R 97	K ATC S S TTT	K CTA' Y Y GCA	A TAC T T GGG	2100
AAG S 204 AAA N 210 ACA H	1 TCA H TTC: S S 1 CTA' Y	CTA Y AGG. G TTT	CTA	rga D CCT L FGA	CTG W TGA' D AGC' A	L TCA Q TAC	C GTA Y TGT V	R TTT L TGT V	GGG. GAA K TGC'	I GTT( L TGG	H GAG R PCT	V AGC A	I TCT L GAT	T CCA Q CAT	GCG R Pep	N GCG R 97	K ATC S S TTT	K CTA' Y Y GCA	A TAC T T GGG	2100
204 AAA N 210 ACA H H 216	1 TCA H TTC: S 1 CTA' Y	CTA Y AGG. G	CCA' H  ACCO P  CTA' Y Y	TGA D CCT' L TGA E E	CTGA W IGA D AGC A	L TCA Q TAC T	C GTA Y TGT V V	TTT L TGT V V	GGG, G K K TGC' A A	I L TGG: G G	H R R PCTC L L	V AGC A CCC P P	I TCT L CGAT I I	CCA Q CAT I	GCG R Pep AAC T T	N GCGG R 97	ATC S S TTT L L	K CTA' Y Y GCA' Q Q	A TAC T T GGG	2100 2160 2220
AAG S 204 AAA N 210 ACA H H 216 ACT	1 TCA H TTC: S 1 CTA' Y	CTA Y AGG. G FTTF F	CCA' H ACC' P CTA' Y Y AAAC' T	rga D CCT' L rga E E	CTGA W I'GA' D AGC' A	L TCAC Q TAC T T	C GTA Y TGT V V	TTT L TGT V V	GGG. GAAA K TGC' A A	I GTTG L IGG G G	H GAGA R PCT( L L	V AGC A	TCT L GAT I	T CCA Q CAT I I	GCG R Pep AAC T T	N GCGG R 97	ATC S S TTT L GAC	K CTA' Y Y GCA' Q Q	A TAC T T GGG G	2100 2160 2220
AAG S 204 AAA N 210 ACA H H 216 ACT L	1 TCA' H 1 TTC. S 1 CTA' Y Y	CTA Y AGG. G FTTF F	CCA' H  ACCO P  CTA' Y Y  AAACO	rga D CCT' L rga E E	CTGA W I'GA' D AGC' A	L TCAC Q TAC T T	C GTA Y TGT V V	TTT L TGT V V	GGG. GAAA K TGC' A A	I GTTG L IGG G G	H GAGA R PCT( L L	V AGC A	TCT L GAT I	T CCA Q CAT I I	GCG R Pep AAC T T	GCG. R 97 CAC	ATC S S TTT L GAC	K CTA' Y Y GCA' Q Q	TAC T T GGG G G	2100 2160 2220
AAG S 204 AAA N 210 ACA H H 216 ACT L	1 TCA H  1 TTC S  1 CTA Y Y TCT L L	CTA Y AGG. G F F F	CCA' H ACC' P CTA' Y Y AAAC' T	rga D CCT' L rga E E	CTGA' D AGC' A CGGA' D	L TCAC Q TAC T T	C GTA Y TGT V V	TTT L TGT V V	GGG. GAAA K TGC' A A	I GTTG L IGG G G	H GAGA R PCT( L L	V AGC A	TCT L GAT I	T CCA Q CAT I I	GCG R Pep AAC T T	GCG. R 97 CAC	ATC S S TTT L GAC	K CTA' Y Y GCA' Q Q	TAC T T GGG G G	2100 2160 2220
210 AAA N 210 ACA H H 216 ACT L 222	1 TCA H 1 TTC. S 1 CTA Y Y 1 TCT' L L	CTA Y AGG. G F F F F	CCA' H ACCO P CTA' Y Y AACO T	TGA D CCT' L TGA E E CGGG	CTGA'  I'GA'  D  AGC'  A  D  D	L TCAC T T CAAC K	C GTA Y TGT V V GAT	R TTT L TGT V V ATT	GGGG GAAA K TGC' A A GCG	I GTT( L IGG' G G	H GAGA R FCT L L TGAA	V AGC A CCC P P AGG	I TCT L GAT I I	T CCCA Q CAT I I TTTT F	P GCG R Pep AAC T T	GCG. R 97 CAC	K ATC S S TTTT L L GAC	K CTA Y Y GCA Q Q TCT	A TAC T T GGG G G	2100 2160 2220
204 AAA N 210 ACA H H 216 ACT L 222 TTA	1 TCA H  1 TTC. S  1 CTA Y Y  L L 1 CAT.	CTA Y AGG. G FTTF F F FGA. E	CCA' H ACCO P CTA' Y Y AAACO T T CAAA	TGAA D CCT' L TGAA E E CGGG	CTGA D AGC' A A GGA' D D	L TCAC T T CAAC K	C GTA Y TGT V V GAT I	R TTT L TGT V V ATT L	GGGG GAAA K TGC' A A GCG	I GTTC L TGGC G G	H GAGA R FCT L L FGA E	V AGC A CCC P P P AGG G	I TCT L CGAT I I CCAT I	T CCAT I I TTTT F	P GCG R Pep AAC T T	N GGCG R 97 CCAC T T T T TGG G	K ATC S S TTTT L L GAC T	K CTA Y Y GCA Q Q TCT	A TAC T T GGG G G S TAG	2100 2160 2220
204 AAA N 210 ACA H H 216 ACT L 222 TTA	1 TCA H  1 TTC. S  1 CTA Y Y  L L 1 CAT.	CTA Y AGG. G FTTF F F FGA. E	CCA' H ACCO P CTA' Y Y AAACO T T CAAA	TGAA D CCT' L TGAA E E CGGG	CTGA D AGC' A A GGA' D D	L TCAC T T CAAC K	C GTA Y TGT V V GAT I	R TTT L TGT V V ATT L	GGGG GAAA K TGC' A A GCGA	I GTTC L TGGC G G	H GAGA R FCT L L FGA E	V AGC A CCC P P P AGG G	I TCT L CGAT I I CCAT I	T CCAT I I TTTT F	P GCG R Pep AAC T T	N GGCG R 97 CCAC T T T T TGG G	K ATC S S TTTT L L GAC T	K CTA' Y Y GCA Q Q TCT L	A TAC T T GGG G G S TAG	2100 2160 2220
204 AAA N 2100 ACA H H 216 ACT L L 222 TTA Y	1 TCA H  1 TTC. S  1 CTA Y Y TCT L L 1 CAT.	CTA Y AGG. G FTTF F F FGA. E	CCA' H ACCO P CTA' Y Y AAACO T T CAAA	TGAA D CCT' L TGAA E E CGGG	CTGA D AGC' A A GGA' D D	L TCAC T T CAAC K	C GTA Y TGT V V GAT I	R TTT L TGT V V ATT L	GGGG GAAA K TGC' A A GCGA	I GTTC L TGGC G G	H GAGA R FCT L L FGA E	V AGC A CCC P P P AGG G	I TCT L CGAT I I CCAT I	T CCAT I I TTTT F	P GCG R Pep AAC T T	N GGCG R 97 CCAC T T T T TGG G	K ATC S S TTTT L L GAC T	K CTA' Y Y GCA Q Q TCT L	A TAC T T GGG G G S TAG	2100 2160 2220 2280
204 AAAA N 210 ACA H H 216 ACT L 222 TTA Y	1 TCA' H  1 TTCA' S 1 CTA' Y Y 1 TCT' L L 1 CAT. I	CTA Y AGG. G F F F E E E ATT.	CCA' H ACCO P CTA' Y Y AAACO T T CAAA	TGA' D CCT' L TGA' E E CGGG G G CAA' N	CTGA W TGA D AGC' A A CTT' F	L TCAC T T CAAC K	C GTA Y TGT V V GAT I	R TTT L TGT V V ATT L TAC	GGG, G GAAA K TGC' A A GCG, R	I GTTGG L IGG G G AAT	H GAGA R FCTC L L FGA E	V AGC A CCC P P AGG G	I TCT L GAT I GCAT I	T CCAT I I TTT F	P GCG R Pep T T CAG S	N GGCG R 97 CCAC T T T T	K ATC S S TTT L L GAC T TGA	K CTA Y Y GCA Q Q TCT L	TAC T T GGG G G G TAG S	2100 2160 2220
204 AAAA N 210 ACA H H 216 ACT L L 2222 TTA Y	1 TCA' H  1 TTCA' S  1 CTA' Y Y 1 TCT' L L 1 CAT' I	CTA Y AGG. G F F F F F ATT. AGGG	CCA' H ACCC P CTA' Y Y AACC T T CAAC	rga D CCT' L rga E E CGG G	CTGA' D AGC' A CGGA' D CTT' F	L TCAC T T CAAC K	C GTA Y TGT V V GAT I GAG S	R TTT L TGT V V ATT L TAC T	GGG, G GAAA K TGC'A A A GCG, T AAAG	I GTTGG L TGGG G G AAT I ACCC	H GAGA R FCT L L FGAA	V AGC P P AGG G TAG	I TCT L CGAT I I CCAT I	T CCA Q Q CAT I I TTT F V CGG	PGCGRPepp	N GGCG, R 977 CCACC T T T GGG G TAAG S	K ATC S S TTTT L L GAC T TGA	K CTA' Y Y GCA Q Q TCT L GGC. A	TAC T T GGG G G S TAG S AAAA K	2100 2160 2220 2280
204 AAAA N 210 ACA H H 216 ACT L L 2222 TTA Y	1 TCA' H  1 TTCA' S  1 CTA' Y Y 1 TCT' L L 1 CAT' I	CTA Y AGG. G F F F F F ATT. AGGG	CCA' H ACCC P CTA' Y Y AACC T T CAAC	rga D CCT' L rga E E CGG G	CTGA' D AGC' A CGGA' D CTT' F	L TCAC T T CAAC K	C GTA Y TGT V V GAT I GAG S	R TTT L TGT V V ATT L TAC T	GGG, G GAAA K TGC'A A A GCG, T AAAG	I GTTGG L TGGG G G AAT I ACCC	H GAGA R FCT L L FGAA	V AGC P P AGG G TAG	I TCT L CGAT I I CCAT I	T CCA Q Q CAT I I TTT F V CGG	PGCGRPepp	N GGCG R 97 CCAC T T T	K ATC S S TTTT L L GAC T TGA	K CTA' Y Y GCA Q Q TCT L GGC. A	TAC T T GGG G G G TAG S	2100 2160 2220 2280
204 AAAA N 2100 ACA H H 2166 ACT L L 2222 TTA Y 2286 A	1 TCA H  1 TTC. S  1 CTA Y Y TCT L L 1 CAT: I GGC: A	CTA Y AGG. G F F F F F ATT. AGGG	CCA' H  ACCO P  CTA' Y Y  AACC T T  CAAC N	rga D CCT' L rga E E CGG G	CTGA' D AGC' A CGGA' D CTT' F	L TCAC T T CAAC K	C GTA Y TGT V V GAT I GAG S	R TTT L TGT V V ATT L TAC T	GGG, G GAAA K TGC'A A A GCG, T AAAG	I GTTGG L TGGG G G AAT I ACCC	H GAGA R FCT L L FGAA	V AGC P P AGG G TAG	I TCT L CGAT I I CCAT I	T CCA Q Q CAT I I TTT F V CGG	PGCGRPepp	N GGCG, R 977 CCACC T T T GGG G TAAG	K ATC S S TTTT L L GAC T TGA	K CTA' Y Y GCA Q Q TCT L GGC. A	TAC T T GGG G G S TAG S AAAA K	2100 2160 2220 2280 2340
204 AAAA N 210 ACA H H 216 ACT L L 222 TTA Y 228 AGC A	1 TCAT H  1 TTC. S  1 CTAT Y Y  1 CAT. I CAT. I	CTA' Y AGG. G FTT' F F GA. E E ATT' F	CCA.  H  ACCO P  CTA.  Y  Y  AACC T  T  CAA.  N	TGAACT T	CTGA' D AGC' A CTT' F	L TCAC T T CAAC K	C GTA Y TGT V V GAT I GAG S	TTTT L TGT V V ATT L TAC T	GGG. GAAG K TGC' A A GCG. R	I GTTGG L TGGG G G AAT I ACC P	H GAGA R FCTCT L L FGAT F	V AGC A CCC P P AGG G TAG	TCT L GAT I I TGAT A CAT A CAT A CAT A	T CCAT I I TTT F V CGG G	GCG R Pep PAAC T T CAG S CGGT V	N GGCG. R 997 CCAC' T T T TGGG S	K ATC S S TTT L L TGA T TGA E	K CTA' Y Y GCA' Q Q TCT L GGC. A	TAC T T GGG G G TAG S AAA K	2100 2160 2220 2280
AAG S 2044 AAA N 2100 ACA H H 2166 ACT L L 2222 TTA Y 2288 AGCC A 2344 AAAA	1 TCAT H  1 TTC. S  1 CTAT Y Y 1 TCT L L 1 CAT I GGCI A	CTA Y AGG. G FFF FFAGG. G	CCA' H ACCO P CTA' Y Y AACCO T T CAAC N	TGAACT T	CTGA W TGA D AGC A A A CGGA D D CTT F	L TCAC Q TAC T T CAAC K TAAC K ACCI	C GTA Y TGT V V GAT I GAGA'	R TTT L TGT V V ATT L TAC T TAC P	GGG. GAAAK K TGC'A A GCG. R	I GTTGG L TGGG G G AATT ACCC P	H GAGA R FCTCT L L FTGAT F	V AGC A CCC P P AGG G TAG S TAG S TAG S TAG S	TCT L  GAT I  TGAT I  TGAA  CGAA  CGAA	T CCAT I I TTT F V CGG G ACT	GCGGR Pepp AACTT CAGGTV AACTT	N GGCG, R P 97 CCAC T T T T T T T T T T T T T T T T T	K ATC S S TTT L L GAC T TGA E	K CTA Y Y GCA Q Q TCT L GGC A TGCC A	TAC T T GGG G G TAG S AAA K TAG R	2100 2160 2220 2280 2340
AAG S 2044 AAA N 2100 ACA H H 2166 ACT L L 2222 TTA Y 2288 AGCC A 2344 AAAA	1 TCAT H  1 TTC. S  1 CTAT Y Y 1 TCT L L 1 CAT I GGCI A	CTA Y AGG. G FFF FFAGG. G	CCA' H ACCO P CTA' Y Y AACCO T T CAAC N	TGAACT T	CTGA W TGA D AGC A A A CGGA D D CTT F	L TCAC Q TAC T T CAAC K TAAC K ACCI	C GTA Y TGT V V GAT I GAGA'	R TTT L TGT V V ATT L TAC T TAC P	GGG. GAAAK K TGC'A A GCG. R	I GTTGG L TGGG G G AATT ACCC P	H GAGA R FCTCT L L FTGAT F	V AGC A CCC P P AGG G TAG S TAG S TAG S TAG S	TCT L  GAT I  TGAT I  TGAA  CGAA  CGAA	T CCAT I I TTT F V CGG G ACT	GCGGR Pepp AACTT CAGGTV AACTT	N GGCG. R 997 CCAC' T T T TGGG S	K ATC S S TTT L L GAC T TGA E	K CTA' Y Y GCA' Q Q TCT L GGC. A	TAC T T GGG G G TAG S AAA K TAG R	2100 2160 2220 2280 2340
2104 AAAA N 2100 ACA H H 2166 ACT L L 222 TTA Y 2288 AGCC A A AAAA K	1 TCA H 1 TTC. S 1 CTA Y Y 1 TCT L L 1 CAT. I GGC. V	CTA Y AGG. G FFF FFAGG. G	CCA' H ACCO P CTA' Y Y AACCO T T CAAC N	TGAACT T	CTGA W TGA D AGC' A A A CGGA D D CTT' F	L TCAC Q TAC T T CAAC K TAAC K ACCI	C GTA Y TGT V V GAT I GAGA'	R TTT L TGT V V ATT L TAC T TAC P	GGG. GAAAK K TGC'A A GCG. R	I GTTGG L TGGG G G AATT ACCC P	H GAGA R FCTCT L L FTGAT F	V AGC A CCC P P AGG G TAG S TAG S TAG S TAG S	TCT L  GAT I  TGAT I  TGAA  CGAA  CGAA	T CCAT I I TTT F V CGG G ACT	GCGGR Pepp AACTT CAGGTV AACTT	N GGCG, R P 97 CCAC T T T T T T T T T T T T T T T T T	K ATC S S TTT L L GAC T TGA E	K CTA Y Y GCA Q Q TCT L GGC A TGCC A	TAC T T GGG G G TAG S AAA K TAG R	2100 2160 2220 2280 2340
204 AAAA N 210 ACA H H 216 ACT TTA Y 228 AGC A A 234 AAAA K	1 TCA' H 1 TTCA' S 1 CTA' Y 1 TCT' L 1 CATA' I GGCA A 1 GGTA' V	CTACY AGG. G FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	CCA.  ACCO P  CTA.  Y  AAACO T  CAAO N  GTA.  Y	TGAACT T	CTGA' D AGC' A CTT' F F F G A CTT' F	L TCAG Q TAC' T T CAAG K ACCI R	C GTA Y TGT V V GAT I GAG S AGA D	R TTT L TGT V V ATT L TAC T TCC. P	GGG, GAAA K TGC' A A GCG, T T AAGG R	I GTTGG L TGGG G G AAAT I ACC P ATTA	H GAGA R FCTCTC L L FGAA E FCTCTC K	V AGC P P AGG G TAG S TCT L	TTCT L CGAT I SCAT I TTGA E	T CCA Q CAT I I TTT F C G G A CT L	P GCG R Pep T T T CAG S GGT V AAC	N GGCG R 977 CCAC T T T T T TGG S TGA D	K ATC S S TTTT L C T TGA T TGA T V TATT I	K CTA' Y Y GCA' Q Q TCT' L GGC. A CCCC' P	TAC T T GGG G G S TAG S AAA K TAG R TGT	2100 2160 2220 2280 2340
2104 AAAA N 2100 ACA H H 2166 ACT L L 222 TTA Y 2288 AGCC A A AAAA K	1 TCAN H 1 TTCC S 1 CTAN Y Y 1 CCAT I 1 GGCI V 1 GGGTI V 1 GGGTI V	CTACY AGGGGGGGGGAATTI	CCA' H ACCO P CTA' Y Y AACC N CAAC N GTA' Y CAAC I CGTA' T CAAC I CGTA' T CAAC I CGTA' T CGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	rgad L rgad E E CGGG G CAAC T	CTG-W IGA-D AGC-A D CTT-F IGA-A AGA-A AGA-A	L TCAG Q TAC' T T CAAG K TAAG R ACCI	C GTA Y TGT V V GAT I GAG S AGA AGA AGG G	R TTTT L TGT V V ATT L TAC T ATC S	GGG. GAAAK K TGC' A A GCG. T AAACC T	I GTTC L IGGG G G AATT I ACCT P ATTI	H GAGA R FCTC L L FGAC F F F F F F F F F F F F F F F F F F F	V AGC P P AGG G TAG S TCT L	TCT L CGAT I I CCAT I CGA E AGC	T CCAT I I TTT F CGG G ACT L CGA	P GCG R Pep AAC T T CAG S GGT V AAC T AAAC AAAC AAAC AAAC AAAC AAAC	N GGCG, R 977 CCAC T T T T T T T T T T T T T T T T T	K ATC S S TTTT L C T TGA T TGA T TGA T T T T T T T T T T T	K CTAY Y Y GCA Q Q TCT L GGC. A CCC: P	TAC T T GGG G G TAG S AAA K TAG R TGT V	2100 2160 2220 2280 2340

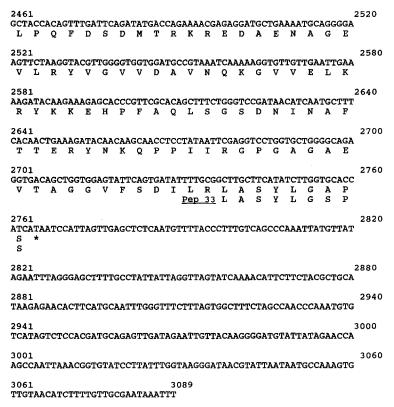


Fig. 1. DNA sequence of carrot of the AK-HSDH cDNA clone and the deduced amino acid sequence. Also shown are the positions of the peptides whose amino acid sequences were determined.

## Isolation of \(\lambda\gt11\) carrot cDNA clones for HSDH

The 1100 bp PCR-amplified DNA fragment was used to screen a carrot cDNA library. The library was constructed in the vector λgt11 and contained cDNA from carrot root poly(A)⁺ RNA. From approximately 250000 plaques four clones were obtained. All of the clones displayed similar restriction digestion patterns and appeared to differ from one another only in length. The complete DNA sequence of the longest clone, HSDH1, was determined. The clone was 2079 bp long and had one long open reading frame starting at bp 1 and reading through bp 1755 for a total of 585 codons.

Clone HSDH1 had no apparent translation start site. RNA blot analysis using the HSDH PCR product as a probe identified a 3000 nucleotide transcript as the main hybridizing band (not

shown). The open reading frame in the clone was not sufficient to code for a protein of  $85 \pm 5$  kDa (the subunit size of the purified carrot enzyme [13]). These facts indicated that clone HSDH1 was lacking ca. 1000 bp at the 5' end.

Isolation of  $\lambda gt11$  carrot cDNA clones for the 5' end of the HSDH gene

In order to enhance the probability of finding HSDH clones with a complete 5' end, we constructed a second cDNA library using oligonucleotide HSDH97 to prime first-strand synthesis. Poly (A)⁺ RNA was isolated from carrot cell suspension cultures. The cDNA was cloned as above into the vector λgt11. 350 000 plaques were screened and 14 positive clones were analyzed. Three of these contained inserts that extended

beyond the 5' end of HSDH1. The longest of these clones, HSDH97-12, was 1009 bp longer at the 5' end than HSDH1.

DNA sequence analysis and comparison to other genes coding for HSDH and AK

The DNA sequence was determined for the total length of clones HSDH1 and HSDH97-12 and for parts of other clones. Figure 1 shows the 3089 bp DNA sequence of the combined HSDH1 and HSDH97-12 clones and the deduced amino acid sequence for the open reading frame from bp 2 through bp 2764. The sequences of four tryptic peptides from the purified carrot HSDH had been determined [24]. These were compared to the deduced amino acid sequence from the cDNA clone. Only one amino acid residue out of 92 did not match ([S] in peptide 90) and this residue was ambiguous in the peptide sequence.

The amino-terminal sequence for carrot HSDH could not be determined from the purified protein. The sequence of the cDNA clone shows a possible translation start site (ATG) at bp 272. If translation were to start at this site, the resulting 831 amino acid protein would have a molecular weight of 90 679 Da, which is in the expected size range.

The deduced amino acid sequence of the carrot HSDH clone was compared to bacterial and fungal HSDH and AK proteins for which there is sequence information available (Table 1). The strongest homology was to the bifunctional E. coli AKI-HSDHI encoded by the thrA gene (Fig. 2). E. coli AKI-HSDHI is 820 amino acids long and can be divided into three general regions. Ca. 248 amino acids at the amino-terminus constitute the AK functional domain; about 359 residues (471– 820) at the carboxy-terminus constitute the HSDH domain, and residues from about 249 to 470 make up a central interface domain. AKII-HSDHII has a similar structure [4, 5]. Proteins with only AK function (E. coli AKIII, Bacillus subtilis AKII, yeast AK) have homology to amino acids 1 through about 500 of AKI-HSDHI. Proteins with only HSDH function (B. subtilis

Table 1. Homology between carrot AK-HSDH and AK and HSDH proteins from other organisms. Comparisons of the carrot AK-HSDK to the three enzymes from E. coli, to the AK of Bacillus subtilis, and to the HSDH from Corynebacterium glutamicum were made using the GAP program of the University of Wisconsin Genetics Computer Group. The comparisons to the B. subtilis HSDH and Saccharomyces cerevisiae AK were based on published alignments between these proteins and the E. coli proteins [16, 17].

	Identity (%)	Homology ¹ (%)	Total identity + homology (%)
E. coli AKI-HSDHI	38	10	48
Amino acids 91-385 ²	30	9	39
Amino acids 386-567	40	9	49
Amino acids 568-921	40	10	50
E. coli AKII-HSDHII	33	9	42
Amino acids 91-385	32	7	39
Amino acids 386-567	19	9	28
Amino acids 568-921	38	10	48
E. coli AKIII (Amino acids 91-385)	28	12	40
B. subtilis AKII (Amino acids 91-558)	33	9	42
B. subtilis HSDH (Amino acids 91-921)	22	8	30
C. glutamicum HSDH (Amino acids 547-921)	25	11	36
S. cerevisiae AK (Amino acids 91-463)	19	8	27

Amino acid replacements: Ile-Val-Leu, Ser-Thr, Lys-Arg, Phe-Tyr, Glu-Asp.

HSDH, Corynebacterium glutamicum HSDH) have homology beginning at about amino acid 500 and extending through the carboxy terminus of AKI-HSDHI. There is homology between the carrot HSDH and E. coli AKI-HSDHI along the full length of the proteins, although it is strongest in the HSDH domain and the interface domain. The homology with AKII-HSDHII is at a simi-

² Amino acid residue numbers refer to the amino acids of the carrot protein as shown in Fig. 2.

Carrot	1 SLSSAISPSSYAAIAAAYSARTPIFNKKKT	AAVLSPLSLFHQSPSLSKTG	50
5	51 IFLHRGRKESSSKFYIAASVTTAVPSLDDS		100
ThrA	1	::      MR.VLKFGGT *****	9
10	O1 CVGSSERIRNVAEIVVEDDSERKL.VVVSA	MSKVTDMMYDLIYKAQSRDD	149
1	:::  : .  : : .:.: .: . :   LO SLANAERFLRVADILESNARQGQVATVLSA * *	. : : : .:   .   :  PAKITNHLVAMIEKTISGQD *	59
15	0 SYESALDAVME.KHKLTAFDLLDED.DLAR		197
é	:: . :: .  . 50 ALPNISDAERIFAELLTGLAAAQPGFPLAQ	:   ::::. :  : ::  LKTFVDQEFAQIKHVLHGIS	109
19	8 IAGHATESFSDFVVGHGELWSAQLLSFVIR		247
11	:  :: :. :::     .::.  : 10 LLGQCPDSINAALICRGEKMSIAIMAGVLE *		151
24	8 GSNQVDPDYLESEKRLEKWFSSNQCQT		291
15	:    .:::::		201
29	2 LKRDGSDFSAAIMGALLRAGQVTIWTDVNG		341
20	:   :   :::     :          LGRNGSDYSAAVLAACLRADCCEIWTDVNG 	VYTCDPRQVPDARLLKSMSY *** *	251
34	2 QEAWEMSYFGANVLHPRTINPVMRYDIPIV		391
25	52 QEAMELSYFGAKVLHPRTITPIAQFQIPCL		298
39	P2 ETEDGLKLESHVKGFATIDNLALINVEGTG		441
29	99 RDEDELPVKGISNLNNMAMFSVSGPG		344
44	12 NVIMISQASSEHSICFAVPESEVKAVAKAL		491
34	5 SVVLITQSSSEYSISFCVPQSDCVRAERAM		394
49	2 NCSILATVGOKMASTPGVSATLFNALAKAN		541
39	95 RLAIISVVGDGLRTLRGISAKFFAALARAN	INIVAIAQGSSERSISVVVN	444
54	2 REDCVRALKAVHSRFYLSRTTIAVGIVGPG		591
44	.: :: .:: .: . .:: .  5 NDDATTGVRVTHQMLFNTDQVIEVFVIGVG 0 0	GVGGALLEQLKRQQSWLK.N	493
59	2 SKIDLRVMGITGSRTMLLSETGIDLSRWRE		641
49	4 KHIDLRVCGVANSKALLTNVHGLNLENWQE	ELAQAKEPFNLGRLIRLVKE	543
64	12 NHF1PSTVIVDCTADSEVASHYHDWLCRGI	HVITPNKKANSGPLDQYLKL	691
	4 YHLL.NPVÍVNČTSSQAVÁDQYADFÉREGFI o	HÝVTPNKKÁNTSSMDYÝHQL o oo	
69	22 RALQRRSYTHYFYEATVVAGLPIITTLQGL		741
59	3 ŘYAAEKŠRRKFLÝDINÝGAĞLÞVÍENLÓNĹ O O	LNAGDELMKFSGILSGSLSÝ o oo o	642
74	2 IFNNFKSTTPFSEVVSEAKAAGYTEPDPRD		791
	3 ÍFGKLDEGMSFSEATRLAREMGYTEPDPRDI O O OO	DLSGMDVARKLLILARETGR OOOO	
79	22 KLELSDIPVQSLVPEPLRGIASAEEFLLQL 		841
	3 ELELADIEIEPVLPAEFNAEGDVAAFMANL	SQLDDLFAARVAKARDEGKV	
	12 LRYVGVVDAVNQKGVVELKRYKKEHPFAQL	::         :	
	3 LRYVGNIDE.DGVCRVKIAEVDGNDPLFKV	KNGENALÁFYSHYYQPLPLV O	791
	22 IRGPGAGAEVTAGGVFSDILRLASY.LGAP		
79	2 LRGYGAĞNDVTAAĞVFADLLRTLSWKLĞV*	. 821	

lar level, except in the interface region, where it is reduced.

All AK amino acid sequences analyzed have the sequence KFGGT near the amino-terminus. This includes the yeast and B. subtilis AKs, E. coli AKIII, and E. coli AKI-HSDHI. (E. coli AKII-HSDHI has the variation KFGGS). The deduced amino acid sequence of the carrot HSDH clone also has this sequence. Another highly conserved region contains the Gly-X-Gly-X-X-Gly motif which is presumed to be part of the NADH/NADPH binding domain of HSDH [16]. This is found in the carrot sequence at amino acids 570–575 as numbered in Fig. 2.

## DNA and RNA blot analysis

All the carrot cDNA clones isolated appeared to represent the same gene. However, it was possible that there were additional genes for AK or HSDH. To test for the existence of other related mRNAs we probed total carrot RNA with the AK-HSDH clone (Fig. 3). Northern blots showed predominantly a message of about 3000 nucleotides. Very minor bands were also seen at about 1500 and 1250 nucleotides. It has not been determined if these are breakdown products of the larger RNA or if they represent second messages. Although Fig. 3 shows a blot probed with a 300 bp fragment from the 5' end of the clone, a similar pattern was seen when the full-length cDNA was used. Little change was seen in signal levels from samples taken at intervals from one to seven days of cell suspension culture growth. On the other hand there was a large difference between the levels observed in root and leaf samples from two-month-old plants; root signal levels were about the same as those of cell culture samples, whereas leaf levels were much reduced. These

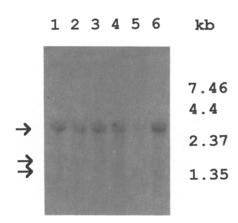


Fig. 3. Northern blot of carrot RNA probed with the AK-HSDH clone. Autoradiograph of RNA expression. Total RNA was isolated from carrot cell suspension cultures and  $10~\mu g$  was loaded in each lane (lane 1, 1 day after transfer; lane 2, 3 days; lane 3, 5 days; lane 4, 7 days); 2-month old leaves (lane 5); and 2-month old roots (lane 6) of total RNA from each sample was separated on an agarose-formaldehydeformamide gel and transferred to nitrocellulose membrane. The blot was hybridized with a  $^{32}P$ -labelled 300 bp Eco RI-Hinc II fragment from the 5' end of the clone.

mRNA levels corresponded with previously observed enzyme activity levels [14].

Genomic carrot DNA was also probed with the AK-HSDH clone (Fig. 4). DNA was digested with several restriction endonucleases. Blots were probed at low stringency and washed at both low and high temperatures. All the blots revealed fairly simple banding patterns, indicating a low number of hybridizing sequences in the genome. The large size of the fragments may indicate the presence of introns as are present in a soybean genomic clone (Gebhardt and Matthews, unpublished). Soybean genomic blots probed with the carrot HSDH clone showed a much more complex pattern (not shown). If there existed related, but only partially homologous, sequences in the carrot genome, higher wash temperatures should have resulted in the elimination of the bands for these sequences.

Fig. 2. Comparison of carrot AK-HSDH and E. coli AKI-HSDH1. The deduced amino acid sequence of the carrot HSDH was compared to the E. coli AKI-HSDHI using the GAP program of the University of Wisconsin Genetics Computer Group. Vertical lines indicate identical amino acids; dots indicate amino acids encoded by similar codons (double dot: codons differing in one nucleotide; single dots: codons differing in two nucleotides). An asterisk (*) under a pair of amino acids indicates identities shared by carrot AK-HSDH, the three E. coli proteins, B. subtilis AKII and yeast AK, a circle (0) indicates identities shared by the carrot protein, E. coli HSDHI and HSDHII, B. subtilis HSDH and C. glutamicum HSDH.

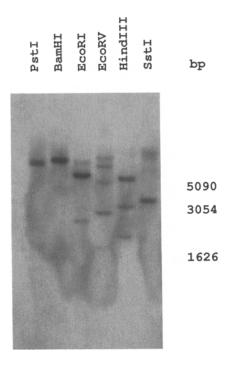


Fig. 4. Southern blot of carrot genomic DNA probed with the AK-HSDH cDNA clone. Carrot leaf DNA was digested with the indicated restriction enzymes, run on an agarose gel and blotted to nitrocellulose membrane. The blot was probed with the full-length AK-HSDH clone.

This was not the case, however, as higher wash temperatures simply reduced the intensity of all bands (not shown).

## Discussion

We have isolated a cDNA clone from carrot that encodes a bifunctional AK-HSDH protein. Two lines of evidence for the bifunctional nature are (1) the strong amino acid homology of the carrot to the two bifunctional *E. coli* proteins, AKI-HSDHI and AKII-HSDHII; (2) the carrot protein (from which the PCR primers were designed) has both AK and HSDH function [24]. There have been no previous reports of a bifunctional AK-HSDH protein in plants. Furthermore, in yeast and in bacterial systems other than *E. coli*, the two enzymes have only been found to exist as separate proteins. Since this is the first AK or

HSDH clone isolated from a plant, it is not known if such fused genes are common in plants. However, we have recently isolated a partial cDNA clone from soybean that has approximately 80% identity with the carrot clone and also appears to encode both AK and HSDH (unpublished).

Although the cDNA clone has a potential start site for translation, it is doubtful that this is the actual start. HSDH and AK have been shown to be localized in the chloroplasts of several plant species [2, 11, 15, 19, 23] and are likely to be located in the chloroplast of carrot. It follows that the carrot AK-HSDH probably has a chloroplast transit peptide at the amino terminus, since all known proteins transported into the chloroplast have been found to possess transit peptides [9]. The reading frame of the clone is continuous from the second base pair through the conserved KFGGT sequence which is found near the amino-terminus of all AK proteins. The 75-80 amino acids encoded by the 5' end of the clone have the character of chloroplast transit peptides: the string is rich in serine, threonine, and small hydrophobic amino acids [9]. The sequence of the amino-terminus of the purified HSDH could not be determined and no precursor with a transit peptide has been identified. Further information is needed to resolve this question.

The deduced amino acid sequence of the carrot AK-HSDH shows homology to bacterial AK and HSDH proteins and to yeast AK. The most striking homology is to E. coli AKI-HSDHI. The overall homology between the carrot HSDH and E. coli AKI-HSDHI (48%) is higher than that between the two E. coli AK-HSDH proteins (46%) [9]. It has been proposed that AKI-HSDHI consists of three domains: an aminoterminal domain with AK activity (amino acids 1 to ca. 248), a carboxy-terminal domain with HSDH activity (amino acids 471 through 820), and an interface domain (amino acids 249–470) [4, 5]. The regions of the carrot protein that are most similar to AKI-HSDHI are the interface and the HSDH domains. In contrast, the homology between the carrot protein and E. coli AKII-HSDHII is lowest in the interface region. The homology in the other regions is only slightly lower

for AKII-HSDHII than for AKI-HSDHI. Both the carrot enzyme and AKI-HSDHI are regulated by threonine, whereas AKII-HSDHII is not [5, 14]. The higher homology between the two proteins may be related to this regulation.

In order to understand the regulation of these genes and the enzyme activities we need to know the number of genes and the number of distinct forms of the enzymes that exist. Studies on the HSDH enzyme of carrot indicate that there is only one HSDH protein, but this protein has different subunit structure and different regulatory properties in vitro depending on solution conditions [13]. Studies on the AK function have shown three forms of enzyme which differ in their sensitivity to lysine and threonine [24]. Some of these forms appear to interconvert. There is no conclusive evidence yet as to whether all the forms of carrot AK and HSDH are separate proteins or if they are forms of the same protein. DNA blots probed with the carrot cDNA indicated a low number of hybridizing sequences in the genome. RNA blots showed predominantly one band of about 3000 nucleotides. In contrast to the carrot results, experiments with a soybean clone show a much more complex pattern of DNA and RNA hybridization (data not shown). It is possible that carrot has only the one AK-HSDH which has both functions and is regulated both by threonine and lysine.

In most bacterial genera separate proteins are responsible for AK and HSDH functions [6, 18]. In these systems there is usually one AK which is regulated by concerted feedback by several end products, most often threonine and lysine. E. coli presents a different scheme in two ways: it has multifunctional AK-HSDH proteins and there is a separate AK for each end product (lysine, threonine, and methionine). Thus far we have identified a single multifunctional AK-HSDH in carrot. If this is the only AK or HSDH in carrot, it would represent a system that combines aspects of the various bacterial systems. It should be interesting to compare AK and HSDH genes from other plants and from blue-green algae to the known genes of carrot and bacteria.

# Acknowledgement

This work was supported in part by the United States Department of Agriculture Office of Competitive Grant (No 87-CRCR-1-2284).

#### References

- 1. Bryan JK: Synthesis of the aspartate family and branched-chain amino acids. In: Miflin BJ (ed) The Biochemistry of Plants: A comprehensive Treatise, vol. 5, pp. 329-357. Academic Press, New York (1980).
- 2. Bryan JK, Lissik EA, Matthews BF: Changes in enzyme regulation during growth of maize. III. Intracellular localization of homoserine dehydrogenase in chloroplasts. Plant Physiol 59: 673-679 (1977).
- Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WI: Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18: 5294– 5299 (1979).
- 4. Cohen GN: The common pathway to lysine, methionine and threonine. In: Herrmann KM, Somerville RL (eds) Amino Acids: Biosynthesis and Genetic Regulation, pp. 147-171. Addison-Wesley, Reading, MA (1983).
- Cohen GN, Saint-Girons I: Biosynthesis of threonine, lysine and methionine. In: Neidhardt FC (ed) Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology, pp. 429–444. American Society for Microbiology, Washington, DC (1987).
- Cohen GN, Stanier RY, Le Bras G: Regulation of the biosynthesis of the amino acids of the aspartate family in coliform bacteria and pseudomonads. J Bact 99: 791–801 (1969).
- Feinberg AP, Vogelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6-13 (1983).
- Kaneko T, Hashimoto T, Kumpaisal R, Yamada Y: Molecular cloning of wheat dihydrodipicolinate synthase. J Biol Chem 265: 17451–17455 (1990).
- Keegstra K, Olsen LJ, Theg SM: Chloroplastic precursors and their transport across the envelope membranes. Annu Rev Plant Physiol Plant Mol Biol 40: 471-501 (1989).
- Keim P, Olson TC, Shoemaker RC: A rapid protocol for isolating soybean DNA. Soybean Genet Newsl 15: 150– 152 (1988).
- 11. Lea PJ, Mills WR, Miffin BJ: The isolation of a lysine-sensitive aspartate kinase from pea leaves and its involvement in homoserine biosynthesis in isolated chloroplasts. FEBS Lett 98: 165–168 (1979).
- 12. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
- 13. Matthews BF, Farrar MJ, Gray AC: Purification and in-

- terconversion of homoserine dehydrogenase from *Daucus carota* cell suspension cultures. Plant Physiol 91: 1569–1574 (1989).
- 14. Matthews BF, Widholm JM: Regulation of lysine and threonine synthesis in carrot cell suspension cultures and whole carrot roots. Planta 141: 315-321 (1978).
- Mills WR, Wilson KG: Effects of lysine, threonine and methionine on light-driven protein synthesis in isolated pea (*Pisum sativum L.*) chloroplasts. Planta 142: 153-160 (1978).
- Parsot C, Cohen GN: Cloning and nucleotide sequence of the *Bacillus subtilis* hom gene coding for homoserine dehydrogenase. J Biol Chem 263: 14654–14660 (1988).
- Rafalski JA, Falco SC: Structure of the yeast HOM gene which encodes aspartokinase. J Biol Chem 263: 2146– 2151 (1988).
- Robert-Gero M, Le Borgne L, Cohen GN: Concerted feedback inhibition of the aspartokinase of *Rhodospirrilum tenue* by threonine and methionine: a novel pattern.
   J Bact 112: 251-258 (1972).
- Sainis JK, Mayne RG, Wallsgrove RM, Lea PJ, Miflin-BJ: Localization and characterization of homoserine dehydrogenase isolated from barley and pea leaves. Planta 152: 491–496 (1981).

- Sanger F, Nicklen S, Coulson AR: DNA sequencing with the chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463-5467 (1977).
- Turano FJ, Jordan RL, Matthews BF: Immunological characterization of in vitro forms of homoserine dehydrogenase from carrot suspension cultures. Plant Physiol 92: 395–400 (1990).
- Turano FJ, Weisemann JM, Matthews BF: Identification and expression of a cDNA clone encoding aspartate aminotransferase in carrot. Plant Physiol 100: 374-381 (1992).
- Wallsgrove RM, Lea PJ, Miffin BJ: Intracellular localization of aspartate kinase and the enzymes of threonine and methionine biosynthesis in green leaves. Plant Physiol 71: 780–784 (1983).
- Wilson BJ, Gray AC, Matthews BF: Bifunctional protein in carrot contains both aspartokinase and homoserine dehydrogenase activities. Plant Physiol 97: 1323-1328 (1991).
- Zakin MM, Duchange N, Ferrara P, Cohen GN: Nucleotide sequence of the metL gene of *Escherichia coli*. J Biol Chem 258: 3028–3031 (1983).